Journal of Computational and Applied Mathematics 309 (2017) 532–541

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Hybrid regression model for near real-time urban water demand forecasting[☆]

Bruno M. Brentan [a], Edevar Luvizotto Jr. [a], Manuel Herrera [b], Joaquín Izquierdo [c,*], Rafael Pérez-García [c]

[a] LHC - FEC, University of Campinas, Campinas, Brazil

[b] EDEn - Department of Architecture and Civil Eng., University of Bath, Bath, UK

[c] Fluing - IMM, Universitat Politècnica de València, Valencia, Spain

ARTICLE INFO

Article history:
Received 7 October 2015
Received in revised form 2 February 2016

MSC:
00-01
99-00

Keywords:
Demand forecasting
Water supply
Fourier series
Support vector regression
Near real-time algorithms

ABSTRACT

The most important factor in planning and operating water distribution systems is satisfying consumer demand. This means continuously providing users with quality water in adequate volumes at reasonable pressure, thus ensuring reliable water distribution. In recent years, the application of statistical, machine learning, and artificial intelligence methodologies has been fostered for water demand forecasting. However, there is still room for improvement; and new challenges regarding on-line predictive models for water demand have appeared. This work proposes applying support vector regression, as one of the currently better machine learning options for short-term water demand forecasting, to build a base prediction. On this model, a Fourier time series process is built to improve the base prediction. This addition produces a tool able to eliminate many of the errors and much of the bias inherent in a fixed regression structure when responding to new incoming time series data. The final hybrid process is validated using demand data from a water utility in Franca, Brazil. Our model, being a near real-time model for water demand, may be directly exploited in water management decision-making processes.

1. Introduction

Long-term water demand forecasting is required for planning and designing new water distribution systems (WDSs) [1]. However, to satisfy safe operation and management of their systems and make right decisions about valve and pump manoeuvres, water utilities need to be acquainted with (near) real-time end-users behaviour regarding water consumption. In addition, having deep knowledge of water demand helps to identify and control possible leakages in the network, when observed consumption and demand prediction diverge far from the expected uncertainty [2,3].

Auto-regressive integrated moving average (ARIMA) based models [4] have been traditionally considered for understanding and modelling urban water demand [5]. ARIMA based models usually treat the problem as a linear correlation among variables and, according to Voitcu and Wong [6], this technique does not always produce predictions with sufficient accuracy, which can harm other processes, such as the control of the system. To cope with this situation, a number of

☆ This work has been partially supported by CAPES Foundation of Brazil's Ministry of Education. The data were provided by SABESP, São Paulo state water management company.

* Corresponding author.
E-mail address: jizquier@upv.es (J. Izquierdo).

http://dx.doi.org/10.1016/j.cam.2016.02.009

data analysis models have been considered more recently. For instance, several authors [7–9] have applied artificial neural network (ANN) architectures to both long and short term demand forecasting. The use of other machine learning tools has also increased during the last years. [10] has performed a comprehensible comparison of various predictive methods for hourly water demand forecasting, suggesting the use of support vector regression (SVR) as one of the models through which it is possible to reach better results. These results are in agreement with the ones previously obtained by Msiza et al. [11], just comparing SVR against ANN architectures for daily water demand prediction. However, the straightforward use of SVR without any internal adaptation when new data arrives (off-line SVR) has two main drawbacks: the computational complexity of the training and validation phases, necessary for parameter tuning, and the selection of the best model among the validated proposals [12,13]. These two phases are highly time-consuming and lead to slow down the whole process, since it is strongly dependent on its associate database size. In addition, off-line predictive models are likely to develop certain growing bias, if models are not updated with the arrival of new data. The model can also become rapidly obsolete in the case of abrupt changes occurring in the forecasting framework. These are the models known as intervened [14], and are a consequence of unexpected changes in the scenario in which the demand is computed. For example, opening and closing valves, extreme variation of weather conditions, appearance of new leaks and celebration of a social events, among others, may change the end-user response regarding water demand.

The majority of hydraulic models proposed in the literature are off-line, addressing a number of purposes such as network design optimization [15], strategy planning [16], and setting optimal pumping schedule to reduce energy [17]. However, off-line models do not represent well the current state of the water supply system for operational purposes, especially in emergency events [18,16]. The new paradigm of on-line modelling in WDSs is a topic of growing interest with the high amount of data information with which water utilities operate nowadays, aiming at making decisions in a very short time [15]. On-line predictive models for water demand forecasting [19,20] emerge to bridge the gap between this constant flow of available information and off-line models, which are not optimized to be updated in near real-time. Through on-line models for water demand, it is possible to improve predictions of water demand and to have better control of such system state variables as flow and pressure [19], by suitable valve operation.

The aim of on-line prediction is to update the current model to a more accurate one, avoiding the computational burden associated with re-calculating the whole process each time new data are available. Vaerenbergh et al. [21] propose a sliding window methodology for kernel regression together with a fast optimization of its associate parameters. Other approaches are based on a fusion of methods for model updating [22]. In this sense, a number of hybrid methodologies for predictive models have been introduced in the literature. These are usually based on ARIMA models combined with alternative models attempting to capture forecasting non-linearities [23]. In general, this fusion of models increases model accuracy and reduces overfitting problems. For example, Aladag et al. [24] proposed a hybrid model linking ARIMA and ANN models; this hybrid model was applied to yearly demand forecast providing accurate results. A similar approach is used in [14] modelling water demand intervened (e.g. by open/close valve manoeuvres). In [25] a hybrid particle swarm optimization - support vector machine (PSO-SVM) based model is built as a predictive model of chemical component concentration in water.

Our working proposal starts by choosing a suitable methodology among the standard machine learning options for regression analysis and run it for short-term water demand forecasting. In this case, we have selected SVR as one of the models that provides better results for water demand forecasting [11,26]. Built on top this model, an on-line process based on Fourier time series is launched to improve the predictions.

The time series error associated with the SVR model is subsequently investigated through and adaptive Fourier series (AFS) technique, enabling to model any possible variation on the original pattern that could arise with the arrival of new data. The combination of these two perspectives endows the model with enough flexibility to be efficiently adapted before any unexpected scenario comes up. Working with a near real-time predictive model for water demand forecasting speeds up decision-making processes on water supply operation and management; in particular, it is useful in water disruption scenarios.

The roadmap for the rest of the paper is as follows: Section 2 describes the hybrid forecasting method where the SVR and AFS equations, together with the hybridization process, are presented. Section 3 presents the water supply case study, the variables involved in the forecasting process, and the evaluation methodology. The results obtained at each step of the method are presented in Section 3.2, where results from SVR and SVR+AFS are analysed separately to give strong support to the hybrid model through clear improvement of results. The efficiency evaluation is also presented in this section. Section 4 presents a discussion on the results and conclusions about this work.

2. Hybrid method: Support vector regression+adaptive Fourier series

This Section introduces our hybrid methodology proposal for near real-time water demand prediction. Firstly, both the off-line model (using SVR) and its corresponding on-line adjustments (via AFS) are succinctly explained. After that, we focus on its combination through modelling the historical record of errors produced by just using the basis off-line model, and justify the use of AFS to adjust those errors. Finally, the statistical parameters for evaluating the accuracy of the proposed model are presented. Also, an efficiency parameter to define a period for retuning the SVR model is developed.